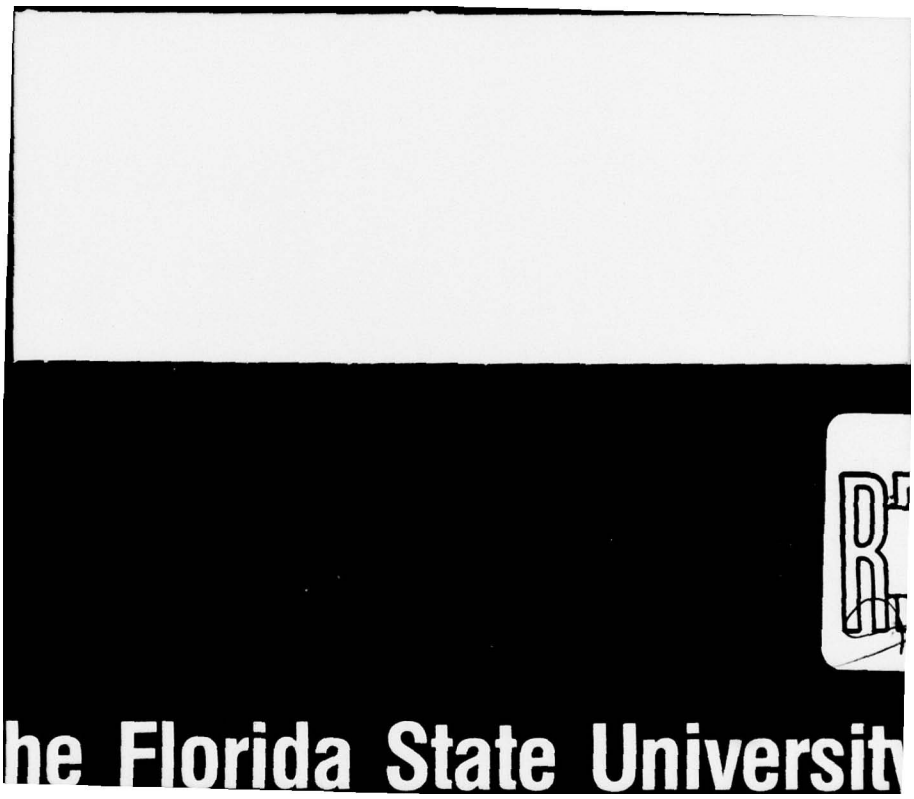
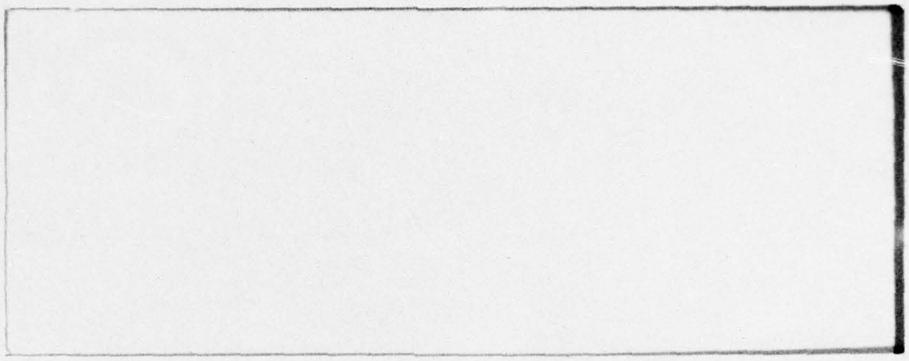
AD-A054 978 FLORIDA STATE UNIV TALLAHASSEE DEPT OF STATISTICS F/G 12/1
TESTING TO DETERMINE THE UNDERLYING DISTRIBUTION USING RANDOMLY--ETC(U)
APR 78 M HOLLANDER, F PROSCHAN AFOSR-76-3109
UNCLASSIFIED FSU-STATISTICS-M458 AFOSR-TR-78-0939 NL

1 OF 1
AD A054978

END
DATE FILMED
7-78
DDC


he Florida State Universit



(2)

TESTING TO DETERMINE THE UNDERLYING DISTRIBUTION USING RANDOMLY CENSORED DATA

Approved for public release;
distribution unlimited.

by

Myles Hollander and Frank Proschan

FSU Statistics Report M458
AFOSR Technical Report No. 6
AFOSR Technical Report No. 79

April, 1978
The Florida State University
Department of Statistics
Tallahassee, Florida 32306

Research sponsored by the Air Force Office of Scientific Research, AFSC, USAF, under Grants AFOSR 76-3109 and AFOSR 74-2581D. The United States Government is authorized to reproduce and distribute reprints for governmental purposes.

Testing to Determine the Underlying Distribution Using Randomly Censored Data

by

Myles Hollander and Frank Proschan

SUMMARY

For right-censored data, we develop a goodness-of-fit procedure for testing whether the underlying distribution is a specified function G. Our test statistic C is the one-sample limit of Efron's (1967) two-sample statistic $\hat{W}$. The test based on C is compared with recently proposed competitors due to Koziol and Green (1976) and Hyde (1977). The comparisons are on the basis of (i) applicability, (ii) the extent to which the censoring distribution can affect the inference, and (iii) power. It is shown that in certain situations the C test compares favourably with the tests of Koziol-Green and Hyde.

Some Key Words: Goodness-of-fit test; Kaplan-Meier estimator; Right-censored data.

1. INTRODUCTION

In the classical non-censored one-sample goodness-of-fit problem, one observes a random sample $X_1,\ldots,X_n$ from a population with distribution function $F(x) = P(X \le x)$; the corresponding survival function is $\bar{F}(x) = P(X > x) = 1 - F(x)$. The null hypothesis asserts that $F(x) = G(x)$, where G is completely specified. The need to generalize this problem to encompass censored data arises because in some situations, such as clinical trials, or life testing, the X's may represent times to the occurrence of an end-point event and the data are usually analyzed before all patients, or items on test, have experienced the event. In the clinical trials context the end-point event could, for example, be relapse, pregnancy, or death. In the life-testing framework, the end-point event could be failure of the inner ring of ball bearings which are on test. In these cases the observations can be viewed as pairs (Z_i, δ_i), $i = 1,\ldots,n$, where

$$Z_i = \min(X_i, T_i), \tag{1.1}$$

$$\delta_i = \begin{cases} 1 \text{ if } Z_i = X_i \ (i^{th} \text{ observation is uncensored}), \\ 0 \text{ if } Z_i = T_i \ (i^{th} \text{ observation is censored}), \end{cases} \tag{1.2}$$

where T_i is the time to censorship of the i^{th} observation. Here we assume that $X_1,\ldots,X_n$ are independent and identically distributed according to a continuous distribution F, $T_1,\ldots,T_n$ are independent and identically distributed according to a continuous censoring distribution $H(x) = P(T \le x)$, and furthermore the T's and the X's are assumed mutually independent. The censoring distribution H is typically, though not necessarily, unknown and is treated as a nuisance parameter.

The goodness-of-fit hypothesis is

$$H_0: \quad F(x) = G(x), \text{ all } x, \tag{1.3}$$

where G is completely specified. Of course, if the censoring distribution H is such that it prevents us from "seeing" F throughout the support of G, we will be unable to use the data to test if H_0 is true. Thus throughout we make the practical assumption that support(G) $\subseteq$ support(H).

In Section 2 we introduce the C statistic, which is the one-sample limit of Efron's two-sample statistic $\hat{W}$. In Section 3 our test of H_0 based on C is compared with competitors due to Koziol and Green (1976) and Hyde (1977), and various advantages and disadvantages of the three procedures are noted. For example, the Koziol-Green test requires the restrictive assumption that the censoring distribution H and the true life distribution F satisfy the relationship given by (3.2), namely that $\bar{H} = \bar{F}^{\beta}$ for some β, $0 < \beta < 2$. However when this condition is indeed satisfied, the Koziol-Green test is consistent against a broader class of alternatives than the C test or Hyde's test. Hyde's test has the disadvantage that the inference to be drawn from the test can be adversely affected by the nuisance parameter H. The C test imposes only a relatively mild restriction on H, namely that the integral in (2.6) converges. Furthermore, the C statistic (2.5) estimates $-\int \bar{G} d\bar{F} (= \int F dG)$, independently of H. The C test is however limited in the sense that it is unable to detect F alternatives to H_0 for which $\int F dG = 1/2$.

Section 3 also contains a Monte Carlo power comparison of the three competitors for (a) normal location alternatives to a hypothesized standard normal, and (b) exponential scale alternatives to a hypothesized exponential with scale parameter 1. In this limited study, the C test compares favourably with the tests of Koziol-Green and Hyde.

Section 4 contains an application of the three goodness-of-fit tests to an updated version of clinical trial data analyzed by Koziol and Green (1976).

2. THE ONE-SAMPLE LIMIT OF EFRON'S TWO-SAMPLE TEST

Efron (1967) has considered the two-sample problem with right-censored data. In addition to the Z's and δ's defined by (1.1) and (1.2), one observes the analogous quantities

$$Z_i' = \min(Y_i, S_i), \tag{2.1}$$

$$\varepsilon_i = \begin{cases} 1 \text{ if } Z_i' = Y_i, \\ 0 \text{ if } Z_i' = S_i, \end{cases} \tag{2.2}$$

where $i = 1,\ldots,m$, and m denotes the size of sample 2. Here $Y_1,\ldots,Y_m$ are independent and identically distributed according to an unknown distribution G, $S_1,\ldots,S_m$ are independent and identically distributed according to $I(x) = P(S \le x)$, and the X's, T's, Y's, and S's are assumed mutually independent.

In the two-sample problem, the null hypothesis is

$$H_0^* : F(x) = G(x), \text{ all } x, \tag{2.3}$$

where the hypothesized common distribution is unspecified. Efron's test of H_0^* is based on the statistic

$$\hat{W} = -\int \hat{\bar{G}}(x) d\hat{\bar{F}}(x), \tag{2.4}$$

where $\hat{\bar{F}}$, $\hat{\bar{G}}$ are the Kaplan-Meier (1958) estimators of $\bar{F}$, $\bar{G}$ respectively. Letting $Z_{(1)} < Z_{(2)} < \ldots < Z_{(n)}$ denote the ordered Z's,

$$\hat{\bar{F}}(x) = \prod_{j=1}^{k-1} \{(n-j)/(n-j+1)\}^{\delta_j}, \quad x \in (Z_{(k-1)}, Z_{(k)}],$$

and $\hat{\bar{F}}(x) = 0$ for $x > Z_{(n)}$. Of course $\hat{\bar{G}}$ is defined analogously.

Invoking an idea due to Moses (1964), we let the sample size m increase without limit. Then the unknown G becomes known, H_0^* reduces to H_0, and $\hat{W}$ becomes

$$C = -\int \bar{G}(x)\, d\hat{\bar{F}}(x). \tag{2.5}$$

The asymptotic mean, variance and distribution of C in this one-sample framework are directly obtainable from Efron's asymptotic results concerning $\hat{W}$. Equivalently, asymptotic properties of C can be derived directly from the result (cf. Efron (1967), Breslow and Crowley (1974), Meier (1975)) that the stochastic process $n^{\frac{1}{2}}\{\hat{\bar{F}}(s) - \bar{F}(s)\}$ tends, as $n \to \infty$, to a Gaussian process with mean 0 and covariance kernel

$$\Gamma(s, t) = -\bar{F}(s)\bar{F}(t) \int_{-\infty}^{s} \{\bar{K}(z)\bar{F}(z)\}^{-1} d\bar{F}(z), \quad s \le t,$$

where $\bar{K}(z) = \bar{F}(z)\bar{H}(z)$. It follows from the continuous mapping theorem that

$$\int n^{\frac{1}{2}}\{\hat{\bar{F}}(x) - \bar{F}(x)\} d\bar{G}(x) \to N(0, \sigma^2),$$

where

$$\sigma^2 = \int\int_{s \le t} \Gamma(s, t)\, d\bar{G}(s)\, d\bar{G}(t) + \int\int_{s > t} \Gamma(s, t)\, d\bar{G}(s)\, d\bar{G}(t)$$

$$= -2 \int_{-\infty}^{\infty} \int_{s}^{\infty} \int_{-\infty}^{s} [\{\bar{F}(s)\bar{F}(t)/\bar{K}(z)\bar{F}(z)\} d\bar{F}(z)\, d\bar{G}(t)\, d\bar{G}(s).$$

Efron shows that, under H_0, σ^2 reduces to

$$\sigma_0^2 = 4^{-1} \int_0^1 z^3 [\bar{K}\{\bar{G}^{-1}(z)\}]^{-1}\, dz = 4^{-1} \int_0^1 z^2 [\bar{H}\{\bar{G}^{-1}(z)\}]^{-1}\, dz, \tag{2.6}$$

and he notes that the integral in (2.6) fails to converge if $\bar{H}(z) = O[\{\bar{G}(z)\}^3]$ as z

approaches 0. If σ_0^2 is finite and if $\hat{\sigma}^2$ is a consistent estimator of σ_0^2, then under H_0,

$$C^* \overset{\text{def.}}{=} n^{\frac{1}{2}}(C - \tfrac{1}{2})/\hat{\sigma} \to N(0, 1).$$

Assuming σ_0^2 is finite, one consistent estimator of σ_0^2 is

$$\hat{\sigma}^2 = 4^{-1} \int_{\overline{G}(Z_{(n)})}^{1} z^3\{\overline{K}_n(\overline{G}^{-1}(z)\}^{-1}dz, \tag{2.7}$$

where $\overline{K}_n$, the empirical survival function of the Z's, is

$$\overline{K}_n(x) = \begin{cases} (n-i+1)/n, & Z_{(i-1)} \le x < Z_{(i)}, \\ 0 & Z_{(n)} \le x, \end{cases}$$

where $Z_{(0)} = -\infty$. Expression (2.7) can be simplified to

$$\hat{\sigma}^2 = 16^{-1} \sum_{i=1}^{n} \{n/(n-i+1)\}[\{\overline{G}(Z_{(i-1)})\}^4 - \{\overline{G}(Z_{(i)})\}^4].$$

Let X be distributed according to G and let X^* be independent of X and have distribution G^*. To test H_0 versus one-sided alternatives $F = G^*$ where $P(X \ge X^*) < \frac{1}{2}$, we reject H_0 if $C^* < -z_\alpha$, and accept H_0 otherwise. To test H_0 versus one-sided alternatives $F = G^*$ where $P(X \ge X^*) > \frac{1}{2}$, we reject H_0 if $C^* > z_\alpha$, and accept H_0 otherwise. Here z_α is the upper α percentile point of a standard normal distribution.

When there is no censoring, our goodness-of-fit test based on C reduces to Moses' (1964) goodness-of-fit test based on the one-sample limit of Wilcoxon's two-sample statistic. That is, with no censoring, $C = \sum_{i=1}^{n} \overline{G}(X_i)/n$ which under H_0 is distribution-free with distribution that of the average of n independent uniform random variables. The test then refers $(12n)^{\frac{1}{2}}(C-\frac{1}{2})$ to the standard normal distribution.

A simplified version of (2.5) for computational purposes is

$$C = \sum_{\{\delta_i=1\}} \bar{G}(Z_{(i)})\hat{f}(Z_{(i)}),$$

where $\hat{f}(Z_{(i)})$, the jump of the Kaplan-Meier distribution at $Z_{(i)}$, is

$$\hat{f}(Z_{(i)}) = \prod_{j=1}^{i-1} \{(n-j+1)/(n-j)\}^{1-\delta_j}$$

at $Z_{(i)}$ uncensored, and at $Z_{(n)}$ uncensored or not.

Although our continuity assumptions preclude ties, in practice if censored observations are tied with uncensored observations, the convention when forming the list of the $Z_{(i)}$'s is to treat the uncensored members of the tie as being less than the censored members of the tie.

3. COMPARISONS OF THE C TEST, THE KOZIOL-GREEN TEST, AND HYDE'S TEST

Competing tests of H_0 have recently been proposed by Koziol and Green (1976) and Hyde (1977).

Koziol-Green (1976) test: Apply the probability integral transformation to the Z's to form new pairs (V_i, δ_i), where $V_i = \min(U_i, L_i)$, $U_i = G(X_i)$, $L_i = G(T_i)$, and δ_i is as before. This reduces the problem to testing whether the distribution of the U's is uniform on (0, 1). The Koziol-Green statistic, a generalization of the Cramér-von Mises statistic to the right-censored situation, is

$$\psi^2 = n \int_0^1 \{\hat{F}_U(t) - t\}^2 dt, \tag{3.1}$$

where $\hat{F}_U$ is the Kaplan-Meier estimator of the distribution of U.

Koziol and Green derive the asymptotic distribution of ψ^2 under the restriction that the censoring distribution H be related to the survival distribution F via

$$\bar{H} = \bar{F}^{\beta}, \tag{3.2}$$

for some β, $0 < \beta < 2$. For this model,

$$P(\delta_i = 0) = \int (1 - \bar{F}^{\beta}) dF = \beta/(\beta+1),$$

so that Koziol and Green interpret β as the censoring parameter. Koziol and Green's asymptotic theory for ψ^2 restricts β to be less than 2, and they give asymptotic critical points of ψ^2 for the models $\beta = 0$ (no censoring) and $\beta = .5, 1$, and 1.5. Thus to implement the ψ^2 test, the user must know β or

estimate β from the data. The estimator $\hat{\beta}$ used by Koziol and Green is obtained by setting

$$\hat{\beta}/(\hat{\beta} + 1) = \sum_{i=1}^{n} (1 - \delta_i)/n \overset{\text{def}}{=} p_c,$$

or

$$\hat{\beta} = p_c/(1 - p_c). \tag{3.3}$$

If the known or estimated value of β is not one of the tabled values, the user can choose between interpolating or deriving new asymptotic percentage points. The condition $\beta < 2$ indicates that the Koziol-Green test will be inappropriate when the expected proportion of censored observations is 2/3 or more. Note that under H_0, the variance expression (2.6), for the special case of model (3.2), reduces to

$$\sigma_0^2 = 4^{-1} \int_0^1 z^{2-\beta} dz,$$

which will converge if $\beta < 3$. This indicates that the C test, in the special setting of model (3.2), can be used when the expected proportion of censored observations is less than 3/4.

Hyde (1977) test: Hyde has generalized the right-censored model to include cases where subject i may enter the study some time ν_i after his lifetime has started.

Let

$$a_i = \log \bar{G}(\nu_i) - \log \bar{G}(Z_i),$$

and set

$$A = \sum_{i=1}^{n} (\delta_i - a_i).$$

Hyde shows that the statistic

$$D = A/\left(\sum_{i=1}^{n} a_i\right)^{\frac{1}{2}} \to N(0, 1),$$

where $\left(\sum_{i=1}^{n} a_i\right)^{\frac{1}{2}}$ is, under H_0, a consistent estimator of the standard deviation of A. Hyde makes the additional assumption that $E(a_i)$ be finite, but that this condition is automatically satisfied follows from Hyde's result that $E(\delta_i - a_i) = 0$. Hyde's statistic, when specialized to our model by setting $\nu_i = 0$ for all i, becomes

$$D = \frac{\sum_{i=1}^{n}\left\{\delta_i + \log \bar{G}(Z_i)\right\}}{\left\{-\sum_{i=1}^{n} \log \bar{G}(Z_i)\right\}^{\frac{1}{2}}}. \tag{3.4}$$

If the failure rate $r(x) = g(x)/\bar{G}(x)$ exists, where $g(x) = (d/dx)G(x)$, then D can be written as

$$D = \frac{\sum_{i=1}^{n}\left\{\delta_i - \int_0^{Z_i} r(u)\,du\right\}}{\left\{\sum_{i=1}^{n} \int_0^{Z_i} r(u)\,du\right\}^{\frac{1}{2}}}.$$

When D is significantly large (small), H_0 is to be rejected in favour of the alternative that the true average failure rate is larger (smaller) than the average failure rate of the hypothesized distribution G.

One disadvantage of Hyde's statistic is that the nuisance parameter H can affect the inference to be drawn. Consider the expectation of the numerator of (3.4). For simplicity we will restrict attention to the numerator since the denominator, under H_0, is a consistent estimator of the standard deviation of the numerator. Furthermore, for ease of calculation, suppose that $\bar{G}(x) = \exp(-x)$. Then

$$E\log\bar{G}(Z_i) = -E\min(X_i, T_i) = -\int\bar{F}(x)\bar{H}(x)dx.$$

Since

$$E\delta_i = \int F(x)dH(x),$$

the mean of the numerator of (3.4) is

$$\Delta(F, H) = \int F(x)dH(x) - \int\bar{F}(x)\bar{H}(x)dx. \tag{3.5}$$

Note also that when F has a density f, we can rewrite (3.5) as

$$\Delta(F, H) = \int\bar{H}(x)\{f(x) - \bar{F}(x)\}dx. \tag{3.6}$$

Suppose now that $F(x) = F_1(x) = 1 - \exp(-x^2)$, the Weibull distribution with shape parameter 2. Then

$$f_1(x) - \bar{F}_1(x) = (2x - 1)\exp(-x^2),$$

so that from (3.6) we see that $\Delta(F_1, H)$ can be made negative by choosing an H which puts most of its probability on $(0, \frac{1}{2})$ and analogously $\Delta(F_1, H)$ can be made positive by choosing an H which puts most of its probability on $(\frac{1}{2}, \infty)$.

To be specific, for

$$H_1(x) = \begin{cases} 1.98x, & 0 \le x \le \tfrac{1}{2}, \\ 1 - \exp(-bx), & \tfrac{1}{2} \le x, \end{cases}$$

with $b = -2\log.01$, we find $\Delta(F_1, H_1) = -.16$. Thus if H_1 is the true censoring distribution, and the sample size n is sufficiently large, Hyde's test will lead to the decision that the failure rate $2x$ of F_1 is "larger" than the constant failure rate 1 of the hypothesized G. However, for

$$H_2(x) = \begin{cases} .0025x, & 0 \le x \le 2, \\ .005 + .99(x-2), & 2 \le x \le 3, \\ 1 - \exp(-cx), & 3 \le x, \end{cases}$$

with $c = -(1/3)\log.005$, we find $\Delta(F_1, H_2) = .11$. Thus when H_2 is the true censoring distribution, and n is sufficiently large, Hyde's test will lead to the decision that the failure rate of F_1 is "smaller" than that of G.

An advantage of the C test proposed in Section 2 is that C estimates $-\int \bar{G}(x)d\bar{F}(x)$, independently of the censoring distribution H. The analogous property for $\hat{W}$ in the two sample situation was a motivating factor in Efron's development of the test based on $\hat{W}$.

In a limited study we have obtained Monte Carlo power comparisons of the tests based on C, D, and ψ^2. Since the tests based on these statistics are only asymptotically exact, our study also provides information about the closeness of the true levels to their nominal asymptotically correct values. Though the tests based on C and D can be one-sided or two-sided, the ψ^2 test is inherently two-sided and thus only two-sided counterparts based on C and D were used. Furthermore, although assumption (3.2) is restrictive and not required by the C or D tests, in fairness to the ψ^2 test we have sampled from situations where (3.2) is satisfied and where β is one of the values for which asymptotic percentage points of ψ^2 are tabled by Koziol and Green.

Table 1 compares the power of the tests based on C, D, and ψ^2 for (a) G specified to be standard normal and F taken to be a normal with a location shift, and (b) G specified to be exponential with scale parameter 1 and F taken to be exponential with a different scale parameter. In each case H was selected so that (3.2) was satisfied with either $\beta = .5$ or $\beta = 1$. The rough indications from Table 1 are that for situation (a) C is to be preferred to ψ^2 which in turn is to be preferred to D. For situation (b) C does best and ψ^2 performs better than D for alternatives close to the null hypothesis but ψ^2 trails D for alternatives more distant from H_0.

Although the values in Table 1 are favourable to the C test, we remind the reader that it is easy to exhibit situations where C will be inadequate. For example, under assumption (3.2) with $0 < \beta < 2$, the ψ^2 test will be consistent when G is specified to be N(0, 1) and F is a scale alternative, $F = N(0, \sigma^2)$ with $\sigma^2 \neq 1$. In such a case the C test will have power remaining approximately at α since for such alternatives $\int F dG = \frac{1}{2}$.

Table 1. Estimated powers of C, D, and ψ^2 tests.

(a) Hypothesized normal, normal location alternatives.

$G(x) = \Phi(x)$, $F(x) = \Phi(x-\theta)$, $\overline{H} = (\overline{F})^{\beta}$

Test	C	D	ψ^2	C	D	ψ^2	C	D	ψ^2	C	D	ψ^2
				n = 20, $\beta = \frac{1}{2}$								
$\alpha\backslash\theta$		0			.25			.5			.75	
.01	.013	.018	.011	.078	.017	.040	.381	.154	.230	.772	.481	.596
.05	.038	.061	.051	.199	.081	.139	.590	.384	.466	.928	.767	.825
.10	.096	.114	.098	.279	.167	.226	.703	.525	.600	.961	.876	.902
				n = 50, $\beta = \frac{1}{2}$								
$\alpha\backslash\theta$		0			.25			.5			.75	
.01	.007	.014	.007	.218	.097	.145	.790	.576	.671	.991	.962	.975
.05	.043	.050	.052	.395	.253	.316	.911	.792	.847	.998	.992	.995
.10	.096	.094	.105	.521	.387	.438	.947	.881	.908	1.000	.997	.999
				n = 20, $\beta = 1$								
$\alpha\backslash\theta$		0			.25			.5			.75	
.01	.018	.010	.013	.056	.011	.025	.346	.107	.158	.726	.403	.417
.05	.044	.060	.057	.152	.058	.103	.527	.329	.351	.872	.688	.680
.10	.086	.127	.103	.227	.128	.163	.628	.462	.453	.918	.825	.798
				n = 50, $\beta = 1$								
$\alpha\backslash\theta$		0			.25			.5			.75	
.01	.010	.013	.017	.182	.063	.097	.743	.514	.521	.981	.921	.927
.05	.052	.057	.054	.353	.219	.255	.884	.780	.762	.994	.980	.978
.10	.097	.108	.109	.445	.328	.359	.926	.861	.845	.997	.992	.988

(b) Hypothesized exponential, exponential scale alternatives.
$\bar{G}(x) = \exp(-x)$, $\bar{F}(x) = \exp(-\theta x)$, $\bar{H} = (\bar{F})^{\beta}$

Test	C	D	ψ^2	C	D	ψ^2	C	D	ψ^2	C	D	ψ^2
				$n = 20, \beta = \frac{1}{2}$								
$\alpha\backslash\theta$		1			.8			.6			.4	
.01	.019	.020	.021	.074	.021	.045	.320	.204	.212	.791	.773	.655
.05	.051	.072	.060	.153	.086	.124	.518	.444	.415	.899	.923	.851
.10	.095	.129	.110	.247	.160	.203	.617	.577	.542	.937	.963	.910
				$n = 50, \beta = \frac{1}{2}$								
$\alpha\backslash\theta$		1			.8			.6			.4	
.01	.012	.011	.014	.145	.075	.096	.666	.620	.537	.996	.996	.992
.05	.044	.052	.052	.312	.236	.261	.827	.805	.775	.999	.999	.998
.10	.089	.105	.099	.417	.348	.360	.886	.886	.845	.999	1.000	.999
				$n = 20, \beta = 1$								
$\alpha\backslash\theta$		1			.8			.6			.4	
.01	.010	.017	.011	.058	.014	.036	.279	.102	.123	.738	.643	.520
.05	.046	.059	.054	.148	.076	.102	.462	.312	.299	.867	.835	.740
.10	.092	.106	.105	.200	.148	.166	.569	.439	.398	.917	.896	.814
				$n = 50, \beta = 1$								
$\alpha\backslash\theta$		1			.8			.6			.4	
.01	.009	.017	.017	.106	.040	.059	.590	.440	.389	.993	.990	.963
.05	.045	.054	.057	.240	.164	.187	.766	.677	.620	1.000	1.000	.990
.10	.093	.102	.111	.335	.259	.273	.834	.784	.733	1.000	1.000	.998

4. EXAMPLE

The data in Table 2, kindly furnished by Drs. J. A. Koziol and S. B. Green, are an updated (March, 1977) version of the data set used by Koziol and Green to illustrate their goodness-of-fit test based on ψ^2. The data correspond to 211 state IV prostate cancer patients treated with estrogen in a Veterans Administration Cooperative Urological Research Group (1967) study. At the March, 1977 closing date there were 90 patients who died of prostate cancer, 105 who died of other diseases, and

16 still alive. Those observations corresponding to deaths due to other causes and those corresponding to the 16 survivors are treated as censored observations (withdrawals). As reported by Koziol and Green (1976), there is a basis for suspecting that had the patients not been treated with estrogen, their survival distribution for deaths from cancer of the prostate would be exponential with mean 100 months. We thus applied the C, D, and ψ^2 statistics to test that the survival distribution is $\bar{G}(x) = \exp(-x/100)$.

For the data of Table 2, $C^* = .69$ with a corresponding two-sided P value of .49. Hyde's statistic is $D = -.17$ with a corresponding two-sided P value of .86. The value of the Koziol-Green statistic for the data of Table 2 is $\psi^2 = 1.02$. Since the proportion p_c of censored observations is $121/211 = .573$, we find from (3.3) that $\hat{\beta} = 1.34$. Entering Table 1 of Koziol and Green (1976) at $\beta = 3/2$ with $\psi^2 = 1.02$ gives $P \approx .14$.

Although all P values are consistent with the hypothesized exponential with mean 100, the value of ψ^2 is more suggestive of a possible deviation from H_0 than are the values of C^* and D. Some insight into this is obtained from Figure 1 which contains plots of the Kaplan-Meier estimator $\hat{\bar{F}}$ and the hypothesized survival function $\bar{G}$.

The visual indication from Figure 1 is consistent with an underlying life distribution F having the property that $\int F dG$ is close to ½. Indeed, the value of C for the data of Table 2 is .51. Recall that C estimates $\int F dG$. Similarly, Figure 1 suggests that the average failure rates of F and G are close and thus it is not surprising that Hyde's statistic assumes a value that is close to its null expected value of zero. However, in a Cramér-von Mises type statistic such as the Koziol-Green ψ^2, the $\hat{F}(x) - G(x)$ differences are squared. Thus the negative deviations, found here mostly for the middle month values, do not "cancel" the positive deviations, found in the early and late months portions of the axis. This is a possible explanation for the relatively lower P value achieved by ψ^2.

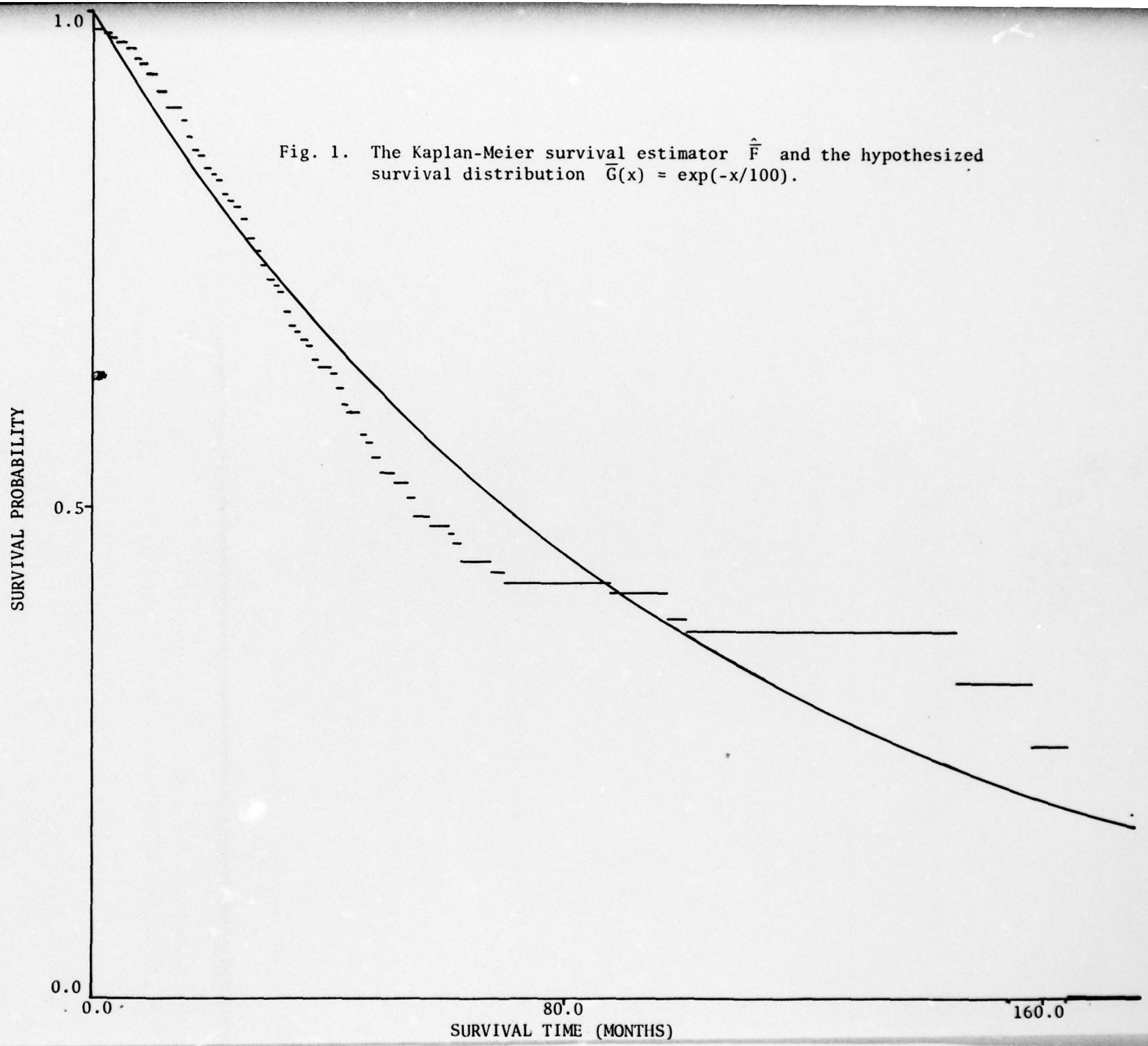


Fig. 1. The Kaplan-Meier survival estimator $\hat{F}$ and the hypothesized survival distribution $\bar{G}(x) = \exp(-x/100)$.

The value of $\psi^2 = 1.02$ was computed for an updated version of the data used by Koziol and Green (1976). Nevertheless, even allowing for updating, our value of 1.02 is not close to their reported value of $\psi^2 = .484$. Although we are unable to obtain the earlier data, we believe the Koziol-Green value is incorrect and suspect that an error in the value reported by Koziol and Green may have arisen through their use, in Appendix 2 of their paper, of the same symbol n to denote both the fixed sample size and the random number of uncensored observations. The possibility of an error of this nature has been confirmed by Drs. Koziol and Green in conversations with the authors of this paper.

Table 2: Survival times and withdrawal times in months for 211 patients (with number of ties given in parentheses)

Survival times: 0(3), 2, 3, 4, 6, 7(2), 8, 9(2), 11(3), 12(3), 15(2), 16(3), 17(2), 18, 19(2), 20, 21, 22(2), 23, 24, 25(2), 26(3), 27(2), 28(2), 29(2), 30, 31, 32(3), 33(2), 34, 35, 36, 37(2), 38, 40, 41(2), 42(2), 43, 45(3), 46, 47(2), 48(2), 51, 53(2), 54(2), 57, 60, 61, 62(2), 67, 69, 87, 97(2), 100, 145, 158.

Withdrawal times: 0(6), 1(5), 2(4), 3(3), 4, 6(5), 7(5), 8, 9(2), 10, 11, 12(3), 13(3), 14(2), 15(2), 16, 17(2), 18(2), 19(3), 21, 23, 25, 27, 28, 31, 32, 34, 35, 37, 38(4), 39(2), 44(3), 46, 47, 48, 49, 50, 53(2), 55, 56, 59, 61, 62, 65, 66(2), 72(2), 74, 78, 79, 81, 89, 93, 99, 102, 104(2), 106, 109, 119(2), 125, 127, 129, 131, 133(2), 135, 136(2), 138, 141, 142, 143, 144, 148, 160, 164(3).

This research was sponsored by the U.S. Air Force Office of Scientific Research. The authors express their gratitude to John Kitchin for programming the Monte Carlo calculations of Section 3, the computations of C, D, and ψ^2 for the data of Table 2, and a computer plot of Figure 1. We also thank James Koziol and Sylvan Green for furnishing the data of Table 2, and John Hyde, George Marsaglia, John McCool, and Jayaram Sethuraman for helpful discussions.

REFERENCES

BRESLOW, N. & CROWLEY, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *Ann. Statist. 2*, 437-453.

EFRON, B. (1967). The two sample problem with censored data. *Proceedings 5th Berkeley Symposium 4*, 831-53.

HYDE, J. (1977). Testing survival under right censoring and left truncation. *Biometrika 64*, 225-30.

KAPLAN, E.L. & MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Am. Statist. Assoc. 53*, 457-81.

KOZIOL, J.A. & GREEN, S.B. (1976). A Cramér-von Mises statistic for randomly censored data. *Biometrika 63*, 465-74.

MEIER, P. (1975). Estimation of a distribution function from incomplete observations. *Perspectives in Probability and Statistics*, Applied Prob. Trust, Sheffield.

MOSES, L.E. (1964). One sample limits of some two-sample rank tests. *J. Am. Statist. Assoc. 59*, 645-51.

VETERANS ADMINISTRATION COOPERATIVE UROLOGICAL RESEARCH GROUP (1967). Treatment and survival of patients with cancer of the prostate. *Surg. Gynecol., and Obstet. 124*, 1011-7.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1. REPORT NUMBER FSU No. M458 AFOSR No. 79 AFOSRN No. 6	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE TESTING TO DETERMINE THE UNDERLYING DISTRIBUTION USING RANDOMLY CENSORED DATA		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER FSU Statistics Report M458
7. AUTHOR(s) Myles Hollander Frank Proschan		8. CONTRACT OR GRANT NUMBER(s) AFOSR 76-3109 AFOSR 74-2581D
9. PERFORMING ORGANIZATION NAME & ADDRESS The Florida State University Department of Statistics Tallahassee, Florida 32306		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME & ADDRESS Air Force Office of Scientific Research Bolling Air Force Base Washington, D.C. 20332		12. REPORT DATE April, 1978 13. NUMBER OF PAGES 18
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS

Goodness-of-fit test; Kaplan-Meier estimator; Right-censored data.

20. ABSTRACT

For right-censored data, we develop a goodness-of-fit procedure for testing whether the underlying distribution is a specified function G. Our test statistic C is the one-sample limit of Efron's (1967) two-sample statistic $\hat{W}$. The test based on C is compared with recently proposed competitors due to Koziol and Green (1976) and Hyde (1977). The comparisons are on the basis of (i) applicability, (ii) the extent to which the censoring distribution can affect the inference, and (iii) power. It is shown that in certain situations the C test compares favourably with the tests of Koziol-Green and Hyde.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: AFOSR-TR-78-0939	2. GOVT ACCESSION NO.
	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): TESTING TO DETERMINE THE UNDERLYING DISTRIBUTION USING RANDOMLY CENSORED DATA	5. TYPE OF REPORT & PERIOD COVERED: Interim Sept.
	6. PERFORMING ORG. REPORT NUMBER: FSU Statistics Report M458
7. AUTHOR(s): Myles Hollander, Frank Proschan	8. CONTRACT OR GRANT NUMBER(s): AFOSR-76-3109
9. PERFORMING ORGANIZATION NAME AND ADDRESS: The Florida State University, Department of Statistics, Tallahassee, Florida 32306	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research /NM, Bolling Air Force Base, Washington, D.C. 20332	12. REPORT DATE: April 1978
	13. NUMBER OF PAGES: 21
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): 23 p.	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

FSU-STATISTICS-M458, TR-6-AFOSR

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Goodness-of-fit test; Kaplan-Meier estimator; Right-censored data.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

For right-censored data, ~~we develop~~ a goodness-of-fit procedure is developed for testing whether the underlying distribution is a specified function G. ~~Our~~ The test statistic C is the one-sample limit of Efron's (1967) two-sample statistic $\hat{W}$. The test based on C is compared with recently proposed competitors due to Koziol and Green (1976) and Hyde (1977). The comparisons are on the basis of ~~(i)~~ applicability, ~~(ii)~~ the extent to which the censoring distribution can affect the inference, and ~~(iii)~~ power. It is shown that in certain situations the C test compares favourably with the tests of Koziol-Green and Hyde.

FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

400 277